

Федеральное агентство по образованию
Государственное образовательное учреждение
высшего профессионального образования
«Казанский государственный технологический университет»
Нижекамский химико-технологический институт

**ИНТЕРНЕТ-ТЕХНОЛОГИИ ПОИСКА
И ПОЛУЧЕНИЯ ДЕЛОВОЙ ИНФОРМАЦИИ**

ТЕКСТЫ ЛЕКЦИЙ

**Нижекамск
2010**

УДК 004.738.5

И 73

Интернет-технологии поиска и получения деловой информации : тексты лекций / сост. О.В. Ибушева, О.В. Руженский, Н.Н. Саримов. – Нижнекамск : Нижнекамский химико-технологический институт (филиал) КГТУ, 2010. – 42 с.

Предназначены для обеспечения курса лекций и подготовки к выполнению практических работ по дисциплине «Интернет-технологии поиска и получения деловой информации» для студентов, обучающихся по специальностям 170500 – Машины и аппараты химических производств, 250100 – Химическая технология органических веществ, 250500 – Химическая технология высокомолекулярных соединений, 250600 – Технология переработки пластических масс и эластомеров, 060800 – Экономика и управление на предприятии (по отраслям) и направлению 552400 – Технология продуктов питания.

Подготовлены на цикле информационных технологий при кафедре АТПП Нижнекамского химико-технологического института.

Печатаются по решению редакционно-издательского совета Нижнекамского химико-технологического института (филиала) КГТУ.

Рецензенты:

Задворнов О.А., доктор физико-математических наук, профессор;

Садьков А.В., кандидат технических наук, доцент.

УДК 004.738.5

© Ибушева О.В., Руженский О.В.,
Саримов Н.Н., 2010.

© Нижнекамский химико-технологический
институт (филиал) КГТУ, 2010.

ТЕМА 1. ОСНОВНЫЕ ПОНЯТИЯ СЕТЕЙ

1. Введение в сети.
2. Принципы построения Internet.
3. Основные сервисы Internet.

1. Введение в сети

В наше время всем понятно, зачем нужен компьютер. На компьютерах активно работают, нередко в офисах стоят десятки компьютеров, на которых по восемь часов в день выполняются какие-то операции. При обилии информации и наличии отдельных компьютеров, на которых она хранится, возникает проблема передачи данных, взаимодействия компьютеров между собой для повышения производительности труда, а также удобства пользователей. При объединении компьютеров одной организации в сеть такая проблема исчезает. **Под сетью в общем случае будем понимать объединение двух и более устройств (например компьютеров) различными средствами связи с целью обмена информацией.**

Локальные сети (LAN)

К локальной можно отнести сеть, объединяющую компьютеры одной организации, если её филиалы не расположены на территории нескольких стран. А вообще, определение «**локальная сеть**» **включает все сети, объединяющие конечное количество пользователей.** Например, сеть района, города, отраслевые сети и т.д.

Локальные сети можно разделить на:

LAN – «чисто» локальные сети (помещение, здание);

CAN – кампусные сети (несколько зданий, например студенческий городок);

MAN – городские сети (город);

WAN – региональные сети (район, область, и т.д.).

Такие сети организуются, исходя из потребностей определенной группы лиц, которым необходимо часто и оперативно обмениваться большим количеством информации.

Глобальные сети (GAN)

Но иногда сети разрастаются настолько, что уже невозможно определить реальное количество их пользователей, которые оказываются разбросанными по всему миру и даже не знают о существо-

вании друг друга. Такие сети, объединяющие локальные сети по всему миру, называются глобальными.

Передача информации в глобальных сетях осуществляется в основном в пакетном режиме, т.е. информация или файл разбиваются на пакеты и передаются независимо друг от друга, а затем снова собираются в исходное сообщение (файл). **Протокол – это набор правил и команд, определяющий передачу информации в сетях.** **Сетевой протокол** определяет передачу на уровне пакетов (низкий уровень), а **прикладной протокол** определяет передачу информации на уровне сервисов (то есть информация более высокого уровня).

Адресация глобальных сетей – это система уникальных имен (адресов), позволяющая передавать информацию. **Сетевая адресация** позволяет нам установить местонахождение компьютера в глобальной сети. **Прикладная адресация** устанавливает местоположение единиц информации. (Примерами единиц информации являются Web-сайт, FTP-сайт, файл и т.д.).

На сегодняшний день наибольшую популярность приобрели две глобальные сети: некоммерческая сеть Fidonet и коммерческая сеть Internet.

2. Принципы построения Internet

Internet берет свое начало с 1969 года с создания системы ARPANET (Advanced Research Projects Net – сеть передовых исследовательских проектов) при министерстве обороны США. Одной из главных целей исследований, проводившихся в ARPANET, было создание сети, сохраняющей работоспособность при нарушении связи между отдельными ее частями. В 80-ые годы Национальный научный фонд США основал сеть (NFSNET), которая была разработана для того, чтобы обеспечить доступ к нескольким суперкомпьютерам для своих главных пользователей, и в 1988 года она заменила сеть ARPANET, став общенациональной сетью на всей территории США.

Физически структуру Internet составляют устройства (и компьютеры в том числе) самых разных типов. Те из них, которые подключены постоянно и участвуют в передаче данных между другими участниками сети, называют **узлами**.

При подключении к Internet компьютер становится участником телекоммуникационного обмена. Вообще, в телекоммуникационном обмене участвуют:

Клиент – это компьютер, программа или человек, включенный в телекоммуникационный обмен – потребитель услуг сети, основной участник телекоммуникационного обмена;

Сервер – это программный комплекс или программно-аппаратная система, предоставляющая удаленный доступ к своим службам или ресурсам с целью обмена информацией. Программы, обрабатывающие запросы клиентов, называются программными серверами (server). А компьютеры, на которых установлены программные сервера, называются физическими серверами;

Узел – это компьютер (или другое техническое устройство), служащий для обеспечения связи между серверами и клиентами.

Таким образом, основа функционирования Internet – работа узлов. Несмотря на то, что многие из узлов аппаратно несовместимы, вся система функционирует надежно благодаря тому, что каждый узел использует стандартные протоколы передачи данных – **TCP/IP** (Transmission Control Protocol/Internet Protocol). **TCP** разбивает данные на пакеты (фиксированного размера в пределах сети) при передаче и соединяет пакеты воедино при приеме. **IP** – система адресации, указывающая, откуда и куда пакет идет.

В общих чертах работу протокола TCP/IP можно описать следующим образом: каждый компьютер, подключенный к Internet, имеет свой уникальный номер, который называется **IP-адресом**. *(Если компьютер имеет несколько сетевых интерфейсов, то и IP-адресов может быть несколько. Например, по одному на каждый интерфейс – сетевую плату.)* Он состоит из 32 битов (4 байта) и записывается обычно как четыре десятичных числа, разделенных точками, например **192.168.22.11** или **217.198.0.34**.

Информация через Internet передается в виде коротких пакетов. То есть, если необходимо переслать длинное сообщение, оно разбивается на множество мелких пакетов, каждый из которых, кроме непосредственно данных, содержит адрес отправителя, адрес получателя и некоторую служебную информацию. Каждый из пакетов передается независимо от других. В некоторых случаях они даже могут следовать разными маршрутами. Когда они приходят к получателю, из них формируется исходное сообщение. Это называется коммутацией пакетов. Итак, протокол TCP/IP регламентирует, как следует разбивать длинное сообщение на пакеты, как должны

быть устроены пакеты, как контролировать прибытие пакетов к месту назначения, что делать в случае ошибок и другие детали.

Таким образом, Internet представляет собой соединенные между собой локальные сети, которые используют протокол TCP/IP и единое пространство IP-адресов. В остальном эти сети независимы административно и финансово. То есть какого-либо единого центра управления Internet не существует.

DNS

Так как человеку не очень удобно запоминать IP-адреса серверов, для того чтобы получить доступ к ним, была разработана система доменных имен (**DNS – Domain Name System**). DNS представляет собой распределённую иерархическую базу данных, которая поддерживается множеством DNS-серверов, каждый из которых отвечает за свою зону. **DNS-сервера преобразуют символьные доменные имена в IP-адреса.** Для пользователя это происходит незаметно. То есть, например, набирая доменное имя web-сервера, пользователь получает доступ к нему, как если бы набрал IP-адрес.

В доменном имени названия доменов принято отделять друг от друга точкой, например **www.support.nchti.ru**, где:

- www – имя компьютера (web-сервер);
- support – название подразделения;
- nchti – название организации;
- ru – географическое положение (страна).

Имя домена, стоящее справа, принято называть доменом первого уровня.

Доменные имена первого уровня это:

- .com (для коммерческих организаций)
- .net (для сетевых организаций)
- .org (для некоммерческих организаций)
- .mil (для военных организаций)
- .gov (для правительственных организаций)
- .edu (для образовательных организаций)

Региональные домены первого уровня:

- .ru (для России)
- .ua (для Украины)
- .us (для США)
- .uk (для Великобритании)

и т.д.

Каждая страна имеет свое доменное имя первого уровня.

Примеры доменных имен второго уровня: **ibm.com**, **php.net**, **gazeta.ru**, **site.ua** и т.д. Примеры доменных имен третьего уровня: **www.tourkz.com**, **send.site.kz**, **smtp.mail.ru**. Существуют и последующие уровни. Однако запоминать очень длинные доменные имена так же сложно, как и IP-адреса.

3. Основные сервисы Internet

Сети Internet интересны не своей способностью передавать информацию вообще, а способностью предоставлять конкретные виды ресурсов (информацию определенного вида или сервис определенного вида), поэтому на базе основного протокола TCP/IP действуют несколько протоколов более высокого уровня, а именно **HTTP**, **FTP**, **SMTP**, **POP3**, **IMAP4** и др.

Электронная почта (E-mail)

E-mail использует протоколы SMTP (для отправки), POP3 или IMAP4 (для получения) электронной корреспонденции. E-mail предусматривает передачу текстовых сообщений от одного пользователя, имеющего электронный адрес, к другому. E-mail может быть использована также для того, чтобы пересылать двоичные файлы, но они должны быть перекодированы в ASCII-формат, поскольку E-mail в Internet может оперировать только информацией в ASCII-формате.

Адрес e-mail имеет следующий формат:

<имя_пользователя>@<имя_домена>

Например: **info@nchti.ru**

World Wide Web

Web была разработана в 1989 году в Европейской лаборатории физики элементарных частиц (CERN) в Женеве (Швейцария).

Web основана на технологии, называемой гипертекстом, которая позволяет организовать текстовую информацию нелинейным способом, используя связи внутри документа и между ними.

Гипертекст – это термин, который впервые был использован американцем Тедом Нельсоном в 1960 году для описания проекта Xanadu – системы публикации, «предназначенной для хранения множества текстов с элементами связи как единого целого и для

обеспечения немедленного доступа к любой его части внутри этого единства».

Язык гипертекстовой разметки HTML (HyperText Markup Language) служит для оформления электронных документов и их связи между собой. Язык состоит из набора различных тегов (указателей). Между тегами вставляется содержимое документа. Теги служат для указания, как выводить содержимое, например цвет, размер, таблицы, вставка рисунков и т.д.

Главным преимуществом HTML перед другими подобными системами является возможность связывать документы между собой. Например, часть одного документа (слово, предложение, рисунок) могут быть ссылкой на другой документ. Это свойство используется для быстрого перехода между документами.

Для того чтобы увидеть исходный код документа – щелкните правой кнопкой мыши в этом окне и выберите в контекстном меню пункт **Просмотр HTML-кода (View page source)**.

HTML является стандартом для электронных документов в Internet. Почти все HTML-документы в сети связаны между собой системой гипертекстовых ссылок.

В сети находится огромное количество серверов, на которых хранится информация в формате HTML. Такие сервера называются Web-серверами. Любой пользователь сети, имея программу для работы с WWW, может получить доступ к подобным ресурсам. Программы для работы с WWW называются браузерами (browser). На сегодняшний день самыми популярными браузерами являются MS Internet Explorer, Netscape Navigator, Opera.

Чтобы открыть определенный документ на удаленном сервере, достаточно иметь доступ к Internet и знать электронный адрес ресурса.

Адреса в WWW. Доступ к ресурсам Internet осуществляется через универсальные локаторы ресурсов (URL). URL состоит из 3-х элементов: названия протокола передачи, имени хоста и путевого имени. Название протокола и имя хоста разделены двоеточием и 2-мя косыми чертами (://). Путь начинается с одной косой черты (/):

<протокол_передачи>://<имя_хоста>/<путевое_имя>

Протокол передачи – http (обычное соединение) или https (безопасное соединение с использованием TLS/SSL). Имя хоста означает либо доменное имя web-узла, либо его IP-адрес и может дополняться указанием порта – адрес, по которому находится сервис на данном сервере. Этот параметр не обязателен. По умолчанию используется порт 80. Путевое имя указывает на конкретный документ на web-узле. Имя файла, в котором хранится ресурс не обязательный параметр. По умолчанию используется имя, которое было указано в настройках Web сервера.

Например: **http://support.nchti.ru:80/proxy.pac**

Некоторые web-серверы в своей работе используют параметры, передаваемые в строке URL. Адрес сервера отделяется от параметров знаком «?», параметры перечисляются через «&»:

<URL>?<имя_парам1=знач1>&<имя_парам2=знач2>...

Обычно пользователю не нужно вводить параметры, они используются в ссылках на web-страницах.

Часто достаточно набрать в адресной строке только имя сервера. На большинстве серверов при этом загружается главная страница. Нужную информацию часто можно найти по ссылкам с главной страницы. Почти во всех современных браузерах имена протоколов (http:// и др.) можно не набирать. Браузер сам определит нужный протокол для данного сервера. То есть в большинстве случаев достаточно набрать адрес (например, www.yandex.ru), чтобы попасть на нужный сайт.

FTP

Для того чтобы передавать файлы по сети Internet, используется стандартный протокол FTP (File Transfer Protocol). FTP-сервера представляют собой файловые архивы. Содержимое ftp-серверов представляет собой часть файловой системы сервера, к которой предоставлен доступ через сеть. Адреса ftp-архивов подчиняются тем же правилам, что и адреса www-сайтов, только указывается другое имя протокола. Например: **ftp://mp3.int.ru**

Обычно для таких серверов необязательно указывать протокол. Браузер может самостоятельно определить необходимый протокол, по которому необходимо подключаться к серверу. Но в случае с адресом в примере все немного сложнее: по этому же адресу ле-

жит и web-сайт, но для доступа используется синтаксис **http://mp3.int.ru**

Так как у протокола http приоритет выше, то если опустить имя протокола и набрать в адресной строке только имя mp3.int.ru, то откроется web-сайт. Для доступа же к ftp-архиву необходимо указать адрес вместе с протоколом. К большинству ftp-серверов возможен анонимный доступ, это значит, что пользователю не требуется вводить имя (login) и пароль, однако некоторые ftp-серверы требуют аутентификации (предоставления пользовательского имени и пароля). Для доступа к таким серверам необходимо знать login и пароль, которые нужно ввести на запрос ftp-клиента (браузера) или использовать url вида:

ftp://<login>:<password>@<доменное_имя_ftp_сервера>

Предпочтение следует отдавать первому способу, т.к. пользовательские регистрационные данные, передаваемые в строке url могут быть легко перехвачены при передаче по сети.

ТЕМА 2. ВВЕДЕНИЕ В ПОИСК ИНФОРМАЦИИ В СЕТИ INTERNET

1. Основная проблема поиска информации.
2. Методы поиска.
3. Этапы поиска.

1. Основная проблема поиска информации

Разберем основную проблему в Internet: необходимо найти некоторую информацию, но неизвестны адреса сайтов, на которых она выложена. Тут на помощь приходят поисковые системы. Обычно для приемлемой работы в Internet достаточно помнить только адреса нескольких поисковых систем. Все остальные адреса можно найти там. Также для поиска используются каталоги ссылок. Там ссылки аккумулируются и разбиваются по темам и подтемам, как в библиотечных каталогах. Последнее время происходит слияние этих двух сервисов и практически на любом поисковом сервере имеется и каталог. Поисковые системы, это обычные сайты, на которые можно зайти с помощью браузера, написать в специальном поле текст запроса и система выдаст адреса документов из своей базы данных, на которых встречаются слова из запроса. На любое одиночное слово поисковая система может выдать несколько тысяч ссылок.

2. Методы поиска

- 1) Обращение по URL (простейший метод поиска). Подразумевает наличие адреса URL и сводится к обращению клиента (программы клиента) к серверу определенного типа, т.е. отправке запроса с использованием определенного протокола.
- 2) Использование набора ссылок. Большинство серверов, представляющие гипертекстовые материалы общего вида, предлагают ссылки и на другие серверы (содержат URL адреса других ресурсов).
- 3) Использование специализированных поисковых механизмов:
 - Поисковые машины. Основная идея поисковых машин – создание базы данных содержащей слова, встречающиеся в документах Internet. В этой базе для каждого слова будет храниться список документов, содержащих это слово.
 - Каталоги ресурсов. В каталогах используется иерархическая (древовидная) модель базы данных, так как любой ресурс,

имеющий URL адрес, описание и другую информацию, подчинен некой классификации.

Поиск ресурсов Internet с использованием поисковых механизмов должен осуществляться по приведенным ниже этапам. Это оптимизирует результат поиска (количество найденных документов будет не столь велико, а соответствие их запросу – более близко к истине).

3. Этапы поиска

1) **Выделение тематических и географических регионов поиска.** Для поиска русских документов лучше использовать русские поисковые механизмы. Это связано с тем, что на их иностранных аналогах эти документы представлены в очень небольших количествах и поиск в них может не дать ожидаемого результата;

2) **Составление тезауруса.** Переход от описания предметной области к формализованным описаниям и построение в конечном счете формального текста, т.е. составление списка ключевых слов;

3) **Выбор поискового инструмента – метода поиска;**

4) **Составление запросов, руководствуясь синтаксисом выбранного поискового инструмента.**

ТЕМА 3. ПОИСКОВАЯ СИСТЕМА ЯНДЕКС

1. Поиск по запросам.
2. Использование формы расширенного поиска.
3. Поиск по каталогу ресурсов.

1. Поиск по запросам

Для поиска на естественном языке или с помощью языка запросов можно воспользоваться простой формой, расположенной по адресу **www.ya.ru**, либо стандартной формой по основному адресу **www.yandex.ru**. Стандартная форма позволяет выбрать область поиска:

- **Везде** – поиск по всем видам ресурсов;
- **Новости** – поиск новостей;
- **Маркет** – поиск товаров;
- **Карты** – поиск населенных пунктов, организаций, адресов, телефонов, измерение расстояний и т.д.;
- **Словари** – поиск информации по словарям и энциклопедиям;

- **Блоги** – поиск по блогам;
- **Картинки** – поиск изображений.

Здесь же расположена ссылка **Все службы**, служащая для перехода к странице, на которой собраны ссылки на все сервисы Яндекса. Поисково-информационные сервисы включают в себя:

- **Поиск** – по всему Интернету;
- **Каталог** – сайты, отобранные вручную;
- **Новости** – последние известия от электронных СМИ;
- **Поиск по блогам и форумам** – что происходит в интернете прямо сейчас;
- **Маркет** – выбор моделей и поиск товаров;
- **Словари** – Даль, Брокгауз, БСЭ, Кругосвет, 5 иностранных языков;
- **Картинки** – поиск изображений;
- **Карты** – масштабируемые карты России и Европы;
- **Адреса** – поиск контактов фирм и организаций;
- **Афиша** – чем заняться в свободное время;
- **Погода** – в России и за рубежом;

– **Телепрограмма** – центральные, спутниковые, региональные каналы.

Трактовка слов в запросе

Независимо от того, в какой форме употреблено слово в запросе, поиск учитывает все его словоформы по правилам русского языка.

*Например, по запросу **танк**, будут найдены ссылки на документы, содержащие **танк, танка, танки, танков, танку** и т.д.*

Если слово набрано в запросе с большой буквы, будут найдены только слова с большой буквы (включая случаи, когда слово первое в предложении). В противном случае будут найдены как слова с большой, так и с маленькой буквы.

*Например, по запросу **Таракан**, будут найдены ссылки на документы, где упоминается название этого индонезийского города, и те случаи упоминания о насекомом, где оно написано с большой буквы (например, с него начинается предложение).*

По умолчанию поиск учитывает все формы заданного слова согласно правилам русского языка. Однако существует возможность поиска по точной словоформе, для этого перед словоформой надо поставить восклицательный знак «!».

*Так по запросу **!Иванову** будут найдены все документы, содержащие словоформу **Иванову**, а по запросу **Иванов** ~~ **!Иванову** – документы, в которых упоминается **Иванов**, кроме тех, которые были найдены по первому запросу (все словоформы, кроме **Иванову**).*

Для того чтобы в результат поиска не попадали страницы, содержащие формы слова совпадающие с другими словами, следует указывать в запросе нормальную форму слова с помощью оператора «!» перед ней.

*Запрос **!!день** найдет все формы слова – **дня, дню, днем** и др., и не найдет форм слова **деть** (одна из которых совпадает со словом **день**).*

Естественно-языковый поиск

Знаки «+» и «-». Если необходимо, чтобы слова из запроса обязательно были найдены, поставьте перед каждым из них знак «+». Для исключения каких-либо слов из результата поиска, перед каждым из них ставится знак «-».

*Например, запрос **мобильный компьютер**, выдаст много ссылок на сайты с разнообразными мобильными устройствами (в*

большинстве случаев речь будет идти о мобильных телефонах). А запрос с «+» **+мобильный +компьютер** найдет документы с упоминанием именно мобильных компьютеров. Если нужно описание Парижа, а не предложения многочисленных турагентств, имеет смысл задать запрос вида **путеводитель по Парижу -агентство -тур**.

Основные операторы

Несколько набранных в запросе слов, разделенных пробелами, означают, что все они должны входить в одно предложение искомого документа. Тот же самый эффект произведет употребление символа «&».

Например, при запросе **мобильный телефон** или **мобильный & телефон**, результатом поиска будет список документов, в которых в одном предложении содержатся и слово **мобильный**, и слово **телефон**. (Эквивалентно запросу **+мобильный +телефон**).

Если же поставить между словами символ «|», то будут найдены ссылки на документы, содержащие любое из этих слов, что удобно при поиске синонимов, взаимозаменяемых слов, близких по смыслу понятий.

Например, при запросе **мобильный | сотовый**, результатом поиска будет список документов, в которых содержатся или слово **мобильный**, или слово **сотовый**.

Еще один знак, тильда «~», позволит найти документы с предложением, содержащим первое слово, но не содержащим второе.

По запросу **телефон ~ сотовый** будут найдены все документы, содержащие слово **телефон**, рядом с которым (в пределах предложения) нет слова **сотовый**.

Для того чтобы подняться от уровня предложения до уровня документа, производится удвоение соответствующего знака. То есть, одинарный оператор «&», «|», «~» ищет в пределах предложения, двойной «&&», «||», «~~» – в пределах документа.

Например, по запросу **производство && (изобутиловый каучук)** будут найдены документы, в которых есть и слово **производство** и словосочетание **изобутиловый каучук**, причем и **изобутиловый каучук**, и **производство** должны быть в одном документе. А запрос **производство изобутилового каучука ~ цена** выдаст все документы со словами **производство изобутилового каучука**, но без слова **цена**.

Поиск с расстоянием

Часто в запросах ищут устойчивые словосочетания. Если поставить их в кавычки, то будут найдены те документы, в которых эти слова идут строго подряд.

*Например, по запросу "красная шапочка" будут найдены документы с этой фразой. При этом контекст **а шапочка у нее была красная** найден не будет.*

Как Яндекс адресует слова. Если все слова в тексте перенумеровать по порядку их следования, то расстояние между словами *a* и *b* – это разница между номерами слов *a* и *b*. Таким образом, расстояние между соседними словами равно 1 (*a* не 0), а расстояние между соседними словами, стоящими в обратном порядке, равно -1. То же самое относится и к предложениям.

Если между двумя словами поставлен знак наклонной черты «/», за которым сразу напечатано число, значит, требуется, чтобы расстояние между ними не превышало этого количества слов.

*Например, по запросу **изготовление /2 деталей**, вы требуете найти документы, в которых содержатся и слово **изготовление**, и слово **деталей**, причем расстояние между ними должно быть не более двух слов и они должны находиться в одном предложении. (Найдутся **изготовление стальных деталей, изготовление деталей машин** и т.д.).*

Если порядок слов и расстояние точно известны, можно воспользоваться указанием типа **/+n**. Так, например, задается поиск слов, стоящих подряд.

*Запрос **вишневый /+1 сад** означает, что слово **сад** должно следовать непосредственно за словом **вишневый**. (К тому же результату приведет запрос "**вишневый сад**").*

В общем виде ограничение по расстоянию задается при помощи пунктуации вида **/(n m)**, где **n** - минимальное, а **m** - максимально допустимое расстояние. Отсюда следует, что запись **/n** эквивалентна **/(-n +n)**, а запись **/+n** эквивалентна **/(+n +n)**.

*Запрос **настройка /(-2 4) компьютера** означает, что между словом **настройка** и словом **компьютера** не более 3 слов (расстояние 4), а между словом **компьютера** и словом **настройка** – не более 1 слова (расстояние 2, обратный порядок слов).*

Практически все знаки можно комбинировать с ограничением расстояния.

Например, результатом поиска по запросу **вакансии ~ /+1 студентов** будут документы, содержащие слово **вакансии**, причем в этих документах слово **студентов** не следует непосредственно за словом **вакансии**.

Когда знаки ограничения по расстоянию стоят после двойных операторов, то употребленные там числа – это расстояние не в словах, а в предложениях.

Запрос **банк && /1 налоги** означает, что слово **налоги** должно находиться в том же самом, либо в соседнем со словом **банк** предложении.

Скобки

Вместо одного слова, в запросе можно подставить целое выражение. Для этого его надо взять в скобки.

Запрос **(история | технология | изготовление) /+1 (сыра | творога)** задает поиск документов, которые содержат любую из фраз **история сыра, технология творога, изготовление сыра, история творога, и т.д.**

Поиск в зонах

Можно искать информацию в «зоне» заголовка (имя «зоны»: **Title**). Синтаксис: **title [поисковое выражение]**.

Запрос **title [НХТИ]** ищет в заголовках документов слово **НХТИ**.

Поиск в определенных элементах

Можно ограничить поиск информации списком серверов или наоборот исключить сервера из поиска (**url**). Для работы не с конкретным URL, а со всеми, начинающимися с данной последовательности символов, используйте **"*"**. Синтаксис: **имя_элемента="имя_ресурса"**.

По запросу **АТПуП ~~ url="www.nchti.ru*"** будут искажаться упоминания специальности **АТПуП** везде, кроме сервера **www.nchti.ru**. Запрос **host="www.nchti.ru"** аналогичен поиску по **url**, но учитываются все зеркала сайта. Запрос **rhost="ru.nchti.www"** аналогичен **host**, но имя хоста записывается в обратном порядке. (Если в конце указано **.***, например **rhost="ru.nchti.*"**, то поиск идет по всем поддоменам заданного домена.) Запрос **like="nchti.ru/file.html"** отбирает для поиска страницы, похожие на заданный URL. Запрос **domain="nchti"** указывает на поиск по страницам, которые расположены в домене **nchti**.

Для поиска по определенному типу файла используется оператор вида: **mime="тип_файла"**.

Например, запрос (**инструкция && синхрофазотрон**) << (**mime="pdf" | mime="rtf"**) будет указывать на поиск только по документам форматов PDF и RTF.

Определены следующие типы файлов:

- HTML,
- PDF (Adobe Portable Document Format),
- DOC (Microsoft Word),
- PPT (Microsoft PowerPoint),
- XLS (Microsoft Excel),
- RTF (Rich Text Format),
- SWF (Macromedia Flash).

Для поиска документов, написанных на определенном языке, используется оператор вида: **lang="язык"**.

Например, запрос **aar** << **lang="de"** будет указывать на поиск документов на немецком языке, содержащих слово **aar**.

Возможен поиск на следующих языках:

- русском (ru),
- украинском (uk),
- белорусском (be),
- английском (en),
- французском (fr),
- немецком (de).

Оператор **date="ГГГГ{*|ММ{*|ДД}}"** указывает, что поиск производится только по страницам, дата которых удовлетворяет заданному условию.

Запрос **date="200310*"** указывает на поиск документов, датированных любым числом октября 2003 года.

Оператор **cat=(ID региона)** или **cat=(ID темы)**. Поиск включает только страницы сайтов, зарегистрированных в каталоге Яндекса, тематическая рубрика или регион которых совпадают с заданным. Идентификатор региона можно получить, прибавив к коду региона смещение 11000000. Идентификатор тематической рубрики получается прибавлением к коду темы смещения 9000000.

Запрос **НХТИ** << **cat=(11000016)** указывает на поиск документов, содержащих слово **НХТИ**, причем сайт зарегистрирован в

каталоге Яндекса как принадлежащий шестнадцатому (Татарстан) региону.

2. Использование формы расширенного поиска

Форма «Расширенный поиск» содержит область ввода запроса (с советом «используйте минус перед словом для его исключения» под областью ввода) и дополнительные элементы, которые помогут тем, кто не владеет языком запросов.

Группа элементов **Искомые слова** содержит:

- 1) расположены относительно друг друга:
 - подряд
 - в одном предложении
 - не очень далеко (по умолчанию)
 - на одной странице
- 2) расположены на странице:
 - где угодно (по умолчанию)
 - в заголовке
- 3) употреблены в тексте:
 - в любой форме
 - точно так, как в запросе

Группа элементов **Страницы** определяет требуемые свойства найденных страниц:

- 1) язык:
 - любой (по умолчанию)
 - русский
 - английский
 - французский
 - немецкий
 - украинский
 - белорусский
- 2) дата:
 - любая (по умолчанию)
 - последние 2 недели
 - последний месяц
 - последние 3 месяца
 - последний год
 - произвольный диапазон

Ниже списка выбора диапазона дат, располагаются элементы ввода дат вручную. Когда возможно, система сама подставляет необходимые числа.

- 1) формат:
 - любой (по умолчанию)
 - HTML
 - PDF (Adobe Acrobat)
 - RTF (Rich Text Format)
 - DOC (Microsoft Word)
 - XLS (Microsoft Excel)
 - PPT (Microsoft PowerPoint)
 - SWF (Macromedia Flash)
- 2) находятся на сайте:
содержит поле для ввода URL сайта.
- 3) регион:
содержит поле для ввода региона (например, **Россия** или **Нижекамск**).
- 4) похожи на страницу:
содержит поле для ввода URL страницы-образа.

Далее имеется элемент для указания числа результирующих ссылок на страницу. По умолчанию на одну страницу выводится 10 ссылок. Можно также установить вывод по 20, 30 и 50 ссылок.

3. Поиск по каталогу ресурсов

Поиск по каталогу заключается в отборе сайтов путем последовательного сужения области поиска в процессе перехода по иерархии подразделов каталога.

ТЕМА 4. ПОИСКОВАЯ СИСТЕМА АПОРТ

1. Поиск по запросам на естественном языке или языке запросов.
2. Поиск по каталогу ресурсов.
3. Использование сервисов АПОРТ.
4. Специальные виды поиска.

1. Поиск по запросам

Для поиска на естественном языке или с помощью языка запросов можно воспользоваться простой формой, расположенной по адресу **www.au.ru**, либо стандартной формой по основному адресу **www.aport.ru**. Стандартная форма позволяет выбрать область поиска:

- Сайты;
- Картинки;
- Новости;
- MP3;
- Рефераты;
- Словари;
- Товары.

Примеры простых запросов

Обычно запрос представляет собой просто одно или несколько слов.

Например, описание специальности АСОИУ. По такому запросу находятся документы, в которых встречаются все слова запроса.

Но существует и неинформативные слова (союзы, предлоги и т.п.), которые в запросе, как правило, игнорируются, так как не несут сами по себе смысловой нагрузки.

Например, по запросу обучение по специальности АСОИУ будут найдены документы, которые одновременно содержат слова «обучение», «специальность», «АСОИУ»; слово «по» будет проигнорировано.

Независимо от того, в какой грамматической форме вы пишете в запросе слово, оно находится в документах во всех своих формах.

Например, по запросу человек шел будут найдены и те документы, которые содержат текст люди идут.

Логические операторы

Оператор логическое «И» (в запросах пишется **И**, **AND**, **&**, **+**) подразумевается, его можно опускать.

*Например, запрос **специальность АТПиП** полностью эквивалентен запросу **специальность И АТПиП**. По любому из этих запросов будут найдены документы, содержащие оба слова.*

Оператор логическое «ИЛИ» (в запросах пишется **ИЛИ**, **OR**, **|**) позволяет искать документы, которые содержат хотя бы одно из слов, указанных в запросе (например, синонимы).

*Например, по запросу **АСОИУ ИЛИ АТПиП** будут найдены документы, содержащие любое из слов запроса, или оба слова одновременно.*

Оператор логическое «НЕ» (в запросах пишется **НЕ**, **NOT**, **~**) позволяет искать документы, которые не содержат слово, указанное в запросе.

*Например, по запросу **работа НЕ дворник** будут найдены документы, содержащие слово **работа**, но не содержащие слово **дворник**.*

Круглые скобки **()** задают порядок действия логических операторов.

*По запросу **курорт ИЛИ санаторий Крым** будут выданы документы, содержащие либо слово **курорт**, либо одновременно слова **санаторий** и **Крым** (оператор **И** действует первым). А по запросу **(курорт ИЛИ санаторий) Крым** будут выданы документы, где встречаются одновременно слова **курорт** и **Крым**, либо **санаторий** и **Крым**.*

Операторы расстояния

Двойные (") или одинарные (') кавычки позволяют находить словосочетание, указанное в них, или близкое к нему (т.е. во-первых, стоп-слова в кавычках игнорируются, как и в обычном запросе; во-вторых, грамматическая форма слов также кавычками не фиксируется).

*По запросу **"яблоки на снегу"** будут найдены документы, содержащие следующие фрагменты: **яблоки на снегу**, **яблоки и снег**, **яблокам под снегом**, **яблоко снег** и т.п.*

Ограничение расстояния в словах (в запросах используются конструкции вида **слN(...)**, **wnN(...)**, **[N,...]**, где **N** – это предел коли-

чества слов) используется, если вы хотите потребовать, чтобы заданные вами слова встречались в пределах N слов.

*По запросу **sl5(лучший курорт Крым)** будут найдены документы, где среди слов **лучший, курорт и Крым** стоит не более двух других слов (то есть общее число слов во фрагменте не более 5). Порядок, в котором встречаются заданные слова, не важен.*

Поиск по адресам

Оператор поиска по адресам (конструкции вида **url=, url:**) позволяет получить список документов, проиндексированных Апортом на заданном сайте.

*Например, по запросу **URL=www.nchti.ru** будут выданы все документы, проиндексированные Апортом на сайте www.nchti.ru.*

Поиск по полям

Как известно, страницы интернет-сайтов содержат не только поле с отображаемой обозревателем информацией (Text – текст), но и различные дополнительные поля:

Title – титул,

Keywords – ключевые слова,

Description – описание).

Кроме того, можно выделить особые элементы и внутри текстового поля:

Alt – комментарии к картинкам.

Anchor – якоря.

Для поиска по различным полям интернет-страниц используются следующие операторы:

title= – указанное после знака равенства слово, или выражение в скобках (знак равенства перед скобками может не ставиться), ищется в заголовках документов.

*Например, по запросу **title=(Нижнекамский институт)** будут искаться документы, в заголовке которых встречаются оба этих слова.*

keywords= – указанное после знака равенства слово, или выражение в скобках (знак равенства перед скобками может не ставиться), ищется в поле META KEYWORDS документа.

*Например, по запросу **keywords(Нижнекамский институт)** будут искаться документы, в которых автор документа указал в ключевых словах оба этих слова.*

description= – указанное после равенства слово или конструкция в круглых скобках должны искажаться в тексте описаний сайтов каталога Апорта и в поле META DESCRIPTION документов.

*По запросу **description=**("Нижнекамский химико-технологический институт") будут найдены корневые документы сайтов, в тексте описаний которых встречается словосочетание **Нижнекамский химико-технологический институт**, а также документы, в которых это словосочетание встречается в поле META DESCRIPTION.*

alt= – указанное после знака равенства слово, или выражение в скобках (знак равенства перед скобками может не ставиться), ищется в полях ALT (комментарии к картинкам) документа.

*Например, по запросу **alt**(Нижнекамский институт) будут искажаться документы, в которых есть изображения с описанием, содержащим оба этих слова.*

anchor= – указанное после равенства слово или конструкция в круглых скобках должны искажаться в тексте ссылок.

*Например, по запросу **anchor=**(скачать музыку) будут искажаться документы, в тексте ссылок на которые встречаются оба этих слова.*

text= – указанное после равенства слово или конструкция в круглых скобках должны искажаться только в обычном тексте. По умолчанию слова запроса ищутся как в тексте, так и во всех указанных выше полях. Чтобы искать только по тексту, выдаваемому в основное окно браузера, следует использовать данный оператор.

*Например, по запросу **text=**(НХТИ ИЛИ КГТУ) будут найдены документы, в которых любое из указанных слов встречается в пределах основного текста документа.*

2. Поиск по каталогу

Поиск по каталогу заключается в отборе сайтов путем последовательного сужения области поиска в процессе перехода по иерархии подразделов каталога.

Начальной страницей **Каталога** является заглавная страница поисковой системы [Апорт](#). На ней размещены ссылки на 15 разделов **Каталога** с гиперссылками на популярные рубрики и алфавитный указатель рубрик. Начальная страница ориентирует пользователя в темах **Каталога** и предоставляет ему возможность выбора направления поиска.

Страница **Каталога** (catalog.aport.ru) содержит следующие элементы:

- 1) Поисковая форма с закладками для выбора области поиска;
- 2) Навигационная строка сервисов Каталога;
- 3) Название рубрики и ее место в дереве Каталога;
- 4) Географические рубрики и количество сайтов по регионам;
- 5) Список подрубрик данной тематической рубрики и количество сайтов в каждой подрубрике;
- 6) Алфавитный указатель тематических рубрик;
- 7) Сортировки описаний сайтов из данной рубрики Каталога;
- 8) Убрать описание – переключатель формы отображения описания;
- 9) Описание сайта.

Поисковая форма служит для ввода поискового запроса. Закладки расширяют или ограничивают границы поиска:

«**Эта рубрика**» – поиск по сайтам данной рубрики;

«**Каталог**» – поиск по сайтам всего Каталога;

«**Интернет**» – поиск по базе всей поисковой системы Апорт.

В каждом случае поиск осуществляется по страницам сайтов с учетом описаний.

Навигационная строка сервисов Каталога содержит ссылки:

«**Главная**» – переход на начальную страницу Апорта;

«**Обзоры**» – обзоры Интернета по темам разделов или рубрик;

«**Популярные сайты**» – список описаний наиболее популярных сайтов разделов Каталога («популярность» определяется по числу переходов на сайт из Каталога);

«**Популярные рубрики**» – список рубрик всего Каталога отсортированных по посещаемости;

«**Новые поступления**» – новые описания сайтов по разделам Каталога за последний день;

«**Все рубрики**» – дерево рубрик и алфавитный указатель тематического Каталога;

«**Добавить сайт**» – форма регистрации сайта в поисковой системе Апорт.

Сортировки описаний сайтов данной рубрики Каталога. Этот механизм сортировок позволяет выбирать наиболее удобный вариант знакомства с содержанием **Каталога**.

- «название» – сортировка аннотаций сайтов по заголовкам в алфавитном порядке;
- «ИЦ» – сортировка аннотаций в зависимости от величины Индекса Цитируемости сайта;
- «дата публикации» – сортировка по дате публикации аннотации в Каталоге.

Убрать описание. Переключатель вида представления описаний дает возможность показывать описания сайтов данной рубрики с аннотациями или без них, а только в виде заголовков.

Описание сайта. Отражает содержание сайта. Дается сайту редактором вручную с учетом данных, представленных владельцем сайта при регистрации. Описание включает:

- заголовок описания сайта и ссылка на сайт;
- аннотация сайта;
- «new» – указатель нового поступления (присутствует некоторое время после публикации описания в Каталоге);
- «i» – дополнительная информация о размещении сайта в Каталоге;
- **** [лига](#) – все сайты поделены на 5 лиг, в зависимости от степени соответствия требованиям, предъявляемым системой АПОРТ;
- [рус, англ, ...] – указатель языков сайта;
- ИЦ – величина индекса цитируемости сайта по версии Апорта;
- Оценка – средняя величина оценок сайта посетителями Каталога и ссылка на форму с отзывами;
- Зеркала – URL зеркал сайта;
- Регион – географический признак сайта.

Дополнительная информация о размещении сайта в Каталоге. Отдельная страница Апорт-каталога содержит:

- заголовок описания и URL сайта;
- дата публикации;
- географический признак;

- список рубрик, в которых размещена ссылка;
- сведения о числе переходов на данный ресурс со страниц Каталога;
- оценка доступности сайта;
- индекс цитируемости.

3. Использование сервисов АПОРТ

АПОРТ предоставляет следующие виды сервисов (проекты):

Апорт-каталог (<http://catalog.aport.ru>). Один из крупнейших в Рунете каталогов интернет-ресурсов. Более 140 000 составленных вручную описаний сайтов, распределенных по 7000 категориям тематического и 3000 географического рубрикаторов. Начальная страница каталога совпадает с главной страницей проекта www.aport.ru.

Апорт-лайт (<http://au.ru>). Страница поиска без рекламы. Эта страница позволяет максимально быстро и эффективно использовать поисковую систему. На странице присутствуют только необходимые для поиска элементы – строка запроса и кнопка. Также, для удобства пользователей существует «подсказка» при наборе запроса в поисковой строке. В качестве «подсказки» используется база запросов, которые были набраны пользователями при обращении к Апорту.

Мапорт (<http://mobile.aport.ru>). Мелодии, логотипы и java-игры для всех популярных моделей мобильных телефонов. Систематизация и навигация позволяет пользователям найти интересующий их контент, а система подписки дает возможность постоянно получать информацию о новинках сайта.

ТВ-программа (<http://tv.aport.ru>). Программа телепередач на текущую и следующую недели. Также можно ознакомиться с анонсами наиболее интересных передач и фильмов.

Погода на Апорте (<http://weather.aport.ru>). Прогноз погоды для различных городов мира.

WAP.Aport.ru (<http://wap.aport.ru>). Версия Апорта для мобильных телефонов.

Апорт-справка (<http://info.aport.ru>). Апорт-справка – это совместный проект с компанией «Евро-Адрес», содержит базу данных справочной информации о более чем 300 000 организаций РФ. С помощью системы поиска, навигации по каталогу товаров и

услуг, региональному рубрикатору пользователь получает список организаций с указанием названия, адреса, телефонов, URL, электронной почты, дополнительных сведений о виде и предмете деятельности. Координаты организаций отображаются на географической карте.

4. Специальные виды поиска

Специальные виды поиска включают в себя:

Поиск новостей – поиск новостной информации.

Поиск mp3-файлов – поиск файлов в форматах *.mp3, *.га и *.mid[i], размещенных на сайтах Рунета. В результатах поиска представляется ссылка на найденный файл и на сайт источник.

Поиск рефератов – поиск рефератов по базе данных сайта Referat.ru.

Поиск знакомств – поиск анкет по базе данных знакомств сайта Омен.

Поиск работы – поиск резюме и вакансий по базе данных сайта Zarplata.ru.

Поиск по картинкам – поиск графических файлов, расположенных на интернет-сайтах.

ТЕМА 5. ПОИСКОВАЯ СИСТЕМА GOOGLE

1. Поиск по запросам на естественном языке или языке запросов.
2. Дополнительные сервисы GOOGLE.
3. Расширенный поиск.

1. Поиск по запросам

Страница поисковой системы GOOGLE на русском языке располагается по адресу www.google.ru. Для поиска информации можно ввести одно или несколько условий поиска в окно поиска (слова или фразу, которые наиболее точно отражают искомую информацию). Дополнительно можно установить переключатель для поиска страниц только на русском языке. После чего нажмите клавишу <Enter> или кнопку <Поиск в Google>. В результате обработки запроса создается страница с результатами, которая представляет собой список веб-страниц связанных со словами введенного запроса, отсортированный по степени релевантности.

Выбор условий поиска. Точность запроса оказывает решающую роль на результаты. Для запроса необходимо отобрать несколько наиболее значимых слов, исключая незначимые. Например, запрос *Нижекамск обучение АСОИУ* окажется удачнее запроса *Я хочу учиться по специальности АСОИУ в Нижнекамске*.

Использование заглавных букв. Поиск в Google выполняется БЕЗ УЧЕТА РЕГИСТРА. Все буквы, независимо от того, как они введены, будут рассматриваться как строчные. Например, дадут одинаковые результаты запросы *Нижекамск*, *НИЖНЕКАМСК* и *НиЖНеКаМСК*.

Автоматическое использование оператора «and» в запросах. По умолчанию Google выдает ссылки на страницы, содержащие все слова запроса. При этом оператор «and» вводить не нужно. Нужно учитывать, что порядок слов в запросе влияет на результаты. Для сужения области поиска необходимо вводить дополнительные условия.

Автоматическое исключение общих слов. По умолчанию Google игнорирует общие слова (например, «где», «и», «когда») и одиночные цифры и буквы, так как они замедляют процесс поиска и мало влияют на качество результатов. Это поведение системы можно изменить с помощью **оператора «+»** или **поиска по словосочетанию**.

Варианты слов (поиск по основе слова). Google в настоящее время использует технологию поиска по основе слова. Везде, где это применимо, Google ведет поиск не только по словам запроса, но и по словам, похожим на некоторые или все указанные в запросе термины. Если в запросе указано *домашний лемур чем кормить*, то Google выполнит поиск и по словам *домашний лемур корм*, а также по другим похожим вариантам искомых слов.

Поиск по словосочетанию. Если результатом поиска должны быть только страницы, содержащие некоторое словосочетание целиком, то нужно использовать кавычки. Поиск по словосочетанию особенно эффективен при поиске имен собственных ("*Александр Невский*"), текстов песен ("*Звезда по имени солнце*") или других известных фраз ("*Быть или не быть*").

Оператор «+». Если общее слово необходимо для получения нужных результатов, то его можно включить в запрос, поставив перед ним знак «+». Например, в запросе *Звездные войны эпизод +I* символ «I» был бы проигнорирован, если бы перед ним не стоял знак «+». Перед знаком «+» обязательно должен быть пробел.

Оператор «-». («Минус-слова»). Иногда необходимо исключить некоторое слово из результатов поиска (например, слово *нота* может относиться как к музыке, так и к политике). Для этого используется знак «-» перед словом. Например, запрос *нота -музыка* указывает на поиск не музыкальных нот, а по всей видимости политических. Обязательно наличие пробела перед знаком «-».

Оператор «OR». Оператор **OR** разделяет слова, из которых хотя бы одно должно присутствовать в результатах поиска. Например, по запросу *Нижнекамск OR Казань* будут найдены страницы со словом *Нижнекамск* или со словом *Казань*.

Поиск по слову и его синонимам. Если перед словом поставить знак «~», то Google должен производить поиск страниц, содержащих слово запроса или его синонимы. Например, по запросу *~auto* Google должен найти страницы со словом *auto*, а также страницы со словами *truck*, *car* и т.д.

Поиск в диапазоне чисел. Если необходимо выполнить поиск, где число запроса может лежать в некотором диапазоне, то можно указать границы этого диапазона, разделив числа двумя точками. Например, запрос *DVD player \$100..150* указывает на поиск DVD-плеера по цене от \$100 до \$500.

«Мне повезёт!». После ввода условий поиска можно нажать на кнопку <Мне повезёт!>, которая сразу откроет самый релевантный сайт по введенному запросу. То есть вообще не отображается страница результатов поиска, но если бы отображалась, то ссылка на этот сайт занимала бы в списке результатов верхнюю позицию. Например, в случае поиска главной страницы Московского университета можно просто ввести *МГУ* и нажать кнопку <Мне повезёт!> вместо <Поиск в Google>. Откроется главная страница сайта www.msu.ru.

Поиск по определенному домену. *АСОИУ site:nchti.ru* – поиск слова *АСОИУ* по всем поддоменам домена *NCHTI.RU*, или *музыка site:ru* – поиск музыки в домене *.RU*.

Поиск одного слова или набора слов, как части URL. *inurl:pdf* – поиск документов по страницам, в адресах которых встречается слово *PDF*; *allinurl:pdf djv* – поиск документов по страницам, в адресах которых встречаются слова *PDF* и *DJV*.

Поиск одного слова или набора слов, как части заголовка web-страницы. *intitle:HXТИ* – поиск документов, в заголовке которых встречается слово *HXТИ*, или *allintitle:HXТИ КГТУ* – поиск документов, в заголовке которых встречаются слова *HXТИ* и *КГТУ*.

Поиск страниц, ссылающихся на URL. *link:nchti.ru* или *link:www.nchti.ru* – поиск страниц ссылающихся на веб-сайт *HXТИ*. Между *link:* и *URL* не должно быть пробела.

Поиск версии страницы, сохраненной в кеше Google. Использование: *cache:www.google.com <слово запроса>*. Например, *cache:www.google.com HXТИ* – поиск страниц со словом *HXТИ* в кеше Google. Данный тип поиска может понадобиться, если например, в данный момент сайт недоступен.

Поиск «похожих» страниц. Для поиска страниц, «аналогичных» некоторой, используются запросы вида *related:www.nchti.ru* – поиск страниц, похожих на страницу *HXТИ*.

Поиск информации о странице (сайте). *info:www.nchti.ru* – поиск информации о странице *www.nchti.ru*.

Поиск определения слова. *define:Компьютер* – поиск определения слова *Компьютер*.

Поиск с исключением порно-сайтов. *safesearch: sex education* – поиск образовательных материалов о сексе, с исключением сайтов, содержащих порно-материалы.

2. Дополнительные сервисы GOOGLE

Рассмотренный выше поиск – это обычный поиск веб-контента. Кроме этого GOOGLE предлагает следующие возможности:

Картинки – поиск картинок.

Карты – поиск адресов, компаний, маршрутов проезда и т.д.

Новости – поиск новостей.

Группы – создание собственных тематических разделов, обсуждение в группах и поиск информации.

Почта Gmail – электронная почта.

ещё> – дополнительные сервисы:

Блоги – поиск информации по блогам.

Каталог – поиск информации в каталоге GOOGLE.

Календарь – собственный календарь.

Фотографии – создание хранилищ фотографий.

Документы – совместная работа над документами.

Вопросы и ответы – здесь пользователи могут задать вопрос или дать ответ на вопросы других пользователей.

Все продукты >> - переход на страницу, на которой собраны ссылки на все виды сервисов GOOGLE.

3. Расширенный поиск

Форма расширенного поиска дает следующие возможности:

1) **Найти результаты:**

– со **всеми** словами;

– с **точной фразой**;

– с **любым** из слов;

– **без** слов.

2) **Язык** – показывать страницы на любом или одном следующих языков:

– английском

– арабском

- армянском
- белорусском
- болгарском
- венгерском
- вьетнамском
- голландском
- греческом
- датском
- иврите
- индонезийском
- исландском
- испанском
- итальянском
- каталанском
- китайском (традиционном)
- китайском (упрощенном)
- корейском
- латышском
- литовском
- немецком
- норвежском
- персидском
- польском
- португальском
- румынском
- русском
- сербском
- словацком
- словенском
- тайском
- турецком
- украинском
- филиппинском
- финском
- французском
- хорватском

- чешском
- шведском
- эсперанто
- эстонском
- японском

3) **Регион** – поиск по всем регионам, либо по сайтам, относящимся к домену определенного государства.

4) **Формат файла** – поиск файлов определенного типа (опция **Только**), либо всех кроме определенного типа (опция **Не**).

Возможные варианты:

- любой формат
- Adobe Acrobat PDF (.pdf)
- Adobe Postscript (.ps)
- Autodesk DWF (.dwf)
- Google Планета Земля KML (.kml)
- Google Планета Земля KMZ (.kmz)
- Microsoft Excel (.xls)
- Microsoft Powerpoint (.ppt)
- Microsoft Word (.doc)
- Rich Text Format (.rtf)
- Shockwave Flash (.swf)

5) **Дата** – показывать веб-страницы, просмотренные впервые:

- в любое время
- за последние 24 часа
- за неделю
- за месяц
- за последние 2 мес.
- за последние 3 мес.
- за последние 6 мес.
- за последний год

6) **Упоминание** – Показывать результаты, содержащие эти слова:

- где угодно на странице;
- в заголовке страницы;
- в основной части страницы;

- в адресе страницы;
 - в ссылках на данную страницу.
- 7) **Домен** – используется для поиска только по определенному домену, либо по всем доменам, кроме указанного.
- 8) **Права использования** – показывать результаты, которые:
- не фильтруются по лицензии;
 - можно бесплатно использовать самому или совместно с другими;
 - можно бесплатно использовать самому или совместно с другими, даже в коммерческих целях;
 - можно бесплатно использовать самому, совместно с другими или изменять;
 - можно бесплатно использовать самому, совместно с другими или изменять, даже в коммерческих целях.
- 9) **SafeSearch** – содержит переключатели:
- Показывать всё.
 - Фильтровать с помощью Безопасного поиска – удаляет из результатов поиска материалы с явно выраженным сексуальным характером.
- 10) Выбор количества результатов, выводимых на страницу – 10, 20, 30, 50 или 100.
- 11) **Поиск по странице:**
- **Похожие** – поиск страниц похожих на указанную;
 - **Ссылки** – поиск страниц, ссылающихся на указанную.
- Здесь же располагается ссылка на страницу **поиска открытого программного кода**. В запросах этой поисковой формы можно использовать специальный синтаксис для поиска нужной информации, например **lang:lisp xml** – поиск программ, написанных на языках LISP или XML. Для более точного поиска используется синтаксис регулярных выражений, с которым можно ознакомиться, перейдя по ссылке <http://www.google.ru/codesearch?hl=ru>.

ТЕМА 6. ПОИСКОВАЯ СИСТЕМА РАМБЛЕР

1. Поиск по запросам на естественном языке или языке запросов.
2. Специальные виды поиска.
3. Использование формы расширенного поиска.

1. Поиск по запросам

Форма поиска на естественном языке или с помощью языка запросов расположена по адресу www.rambler.ru. Стандартная форма позволяет выбрать область поиска:

Интернет – поиск по всем видам ресурсов;

Новости – поиск новостей;

Картинки – поиск изображений;

Покупки – поиск товаров;

Топ100 – поиск информации в рейтинге Топ100;

Словари – поиск информации по словарям и энциклопедиям.

Операторы. Поиск операторов в документе не производится, они служат лишь инструкцией поисковой машине. Все операторы поисковой машины бинарные, то есть имеют левую и правую часть, каждая из которых также является запросом (по умолчанию состоящим из одного слова). Для изменения сферы действия операторов (группировки нескольких слов запроса в аргумент оператора) применяются скобки и кавычки.

Оператор && (логическое И). Два запроса, соединенные оператором &&, образуют сложный запрос, которому удовлетворяют только те документы, которые одновременно удовлетворяют обоим этим запросам. Иными словами, по запросу **собака && кошка** найдутся только те документы, которые содержат и слово «собака», и слово «кошка».

Но слово «собака» и слово «кошка» на найденной странице могут находиться в самых разнообразных местах, как рядом – в одном предложении, так и в разных предложениях, и даже разных статьях. Для того чтобы дать понять поисковой машине, что слова должны находиться близко друг к другу, можно использовать **мо-**

дифицированное И – &, для управления им служат регулирующие операторы > и <. Чтобы расстояние между словами в результате поиска было меньше заданного по умолчанию, можно использовать конструкцию &< или &<<, чем больше регулирующих операторов, тем сильнее уменьшается расстояние. Чтобы увеличить исходное расстояние, нужно применить обратный оператор: &> или &>>.

Оператор && не имеет степеней регулировки и является оператором **И**, при котором в запрос попадают даже самые далеко стоящие друг от друга слова.

Например, сравните результат поиска запроса **остап почувствовал, что его понесло & 12 стульев** и **остап почувствовал, что его понесло && 12 стульев**. В первом случае будут преобладать документы с указанной цитатой из произведения «Двенадцать стульев», а во втором случае поисковая машина найдет текст самого произведения или его большой фрагмент.

Оператор || (логическое ИЛИ). Сложному запросу, состоящему из двух запросов, соединенных оператором ||, удовлетворяют все документы, удовлетворяющие хотя бы одному из этих двух запросов. По запросу **собака || кошка** найдутся документы, в которых есть хотя бы одно из двух слов – слово «собака» или слово «кошка» (или оба эти слова вместе).

Существует модификация оператора **ИЛИ** – |. Ее обычно используют для корректировки выдаваемого результата по запросу путем добавления к запросу необязательных, но возможно встречающихся на искомым страницах слов. Обычно их применяют, когда количество найденных по запросу страниц довольно велико и необходимо повысить рейтинг искомым страниц, чтобы они были выведены в числе первых.

Оператор NOT (логическое НЕ). Оператор **NOT** позволяет сформировать запрос, которому отвечают документы, удовлетворяющие левой части запроса и не удовлетворяющие правой. Так, результатом поиска по запросу **собака NOT кошка** будут все документы, в которых есть слово «собака» и нет слова «кошка».

Это особенно полезно в случаях, если искомая словоформа является одновременно формой другого слова: **Женя** – имя собственное, а также деепричастие. Если мы ищем именно человека по имени Женя, в запросе можно написать **Женя NOT женить**.

Если оператор явно не указан, поисковая машина сама расставляет операторы так, как считает нужным. Так, запрос **регистрация доменов** будет истолкован как **регистрация && доменов**.

Запрос из нескольких слов, перемежающихся операторами, будет истолкован в соответствии с их приоритетом. Оператор **NOT** имеет самый высокий приоритет, следующий по приоритету оператор **&& (И)**, поэтому запрос из нескольких слов при обработке сначала группируется по операторам **NOT, && (И)** и лишь потом по операторам **|| (ИЛИ)**. Изменить порядок группировки можно использованием скобок.

Существует модификация оператора **И** – оператор **&&&**. Он применяется для расширения границ поиска. В этом случае поиск будет производиться не в пределах одного документа, а в пределах одного сайта.

Следующая группа специальных операторов поможет определить приоритет слов в поисковом запросе. Эти операторы должны стоять непосредственно перед словами поиска (без пробела) или перед кавычками, которыми окружена группа слов:

- Знак «+» – означает, что это слово обязательно должно быть найдено,
- Знак «?» – означает, что могут быть также найдены документы без этого слова.

Например, в запросе **+телефонная база ?Москвы** слово «Москва» будет не обязательным, а слово «телефонная» – необходимым.

Кавычки. Для поиска цитат можно использовать двойные кавычки. Слова запроса, заключенного в двойные кавычки, ищутся в документах именно в том порядке и в тех формах, в которых они встретились в запросе.

Таким образом, двойные кавычки можно использовать и просто для поиска слова в заданной форме (по умолчанию слова находятся во всех формах). Например, запросу **самолет "заправился" посадка** удовлетворяет документ, содержащий текст **самолет совершил посадку и заправился**, и не удовлетворяет документ, содержащий **самолет совершил посадку, чтобы заправиться**.

Другой пример – использование кавычек с ужесточающим поиском оператором **+**. При поиске **"что? где? когда?"** можно найти **что? где? когда?** и **что где когда**. Для ужесточения критерия по-

иска можно добавить к запросу +: +"**что? где? когда?**". В таком случае в результате будут найдены только страницы с **что? где? когда?**, другое написание будет недопустимо.

Скобки. При построении запросов может возникать необходимость объединения слов запроса в группы, которые будут аргументами некоторого оператора. Такие группы заключаются в скобки.

Часть запроса, заключенная в скобки, сама является запросом, и на нее распространяются правила языка построения запросов. Использование скобок позволяет строить вложенные запросы и передавать их операторам в качестве аргументов, а также перекрывать приоритеты операторов, принятые по умолчанию.

Если запрос без скобок **машина самолет || аэродром** означает *найти документы, содержащие либо слова «машина» и «самолет», либо слово «аэродром»*, то запрос со скобками **машина (самолет || аэродром)** означает *найти документы, содержащие слово «машина» и одно из слов «самолет» или «аэродром»*.

2. Специальные виды поиска

Специальные (дополнительные) сервисы Рамблер включают в себя:

Новости – поиск новостей;

Картинки – поиск картинок;

Покупки – поиск;

Топ100 – поиск в материалах сайтов Топ100;

Словари – английский и немецкий словари;

Карты – карты российских городов;

Рамблер ТВ – поиск в программах телепередач;

Профессионалам – поиск с использованием специальных поисковых машин:

– [поиск по библиотеке MSDN library](#);

– поиск по архиву почтовых рассылок FreeBSD;

– поиск по архиву списка рассылки Linux Kernel.

Лекарства – поиск лекарств в аптеках Москвы;

Рамблер-Право – поиск ответов на юридические вопросы.

3. Использование формы расширенного поиска

Форма расширенного поиска предоставляет следующие возможности:

- 1) **Поиск по тексту:**
 - всего документа;
 - названия (title);
 - гиперссылок.
 - 2) **Искать слова запроса:**
 - все («и»);
 - хотя бы одно («или»);
 - точную фразу.
 - 3) **Расстояние между словами запроса:**
 - не ограничивать;
 - ограничивать.
 - 4) **Исключить документы, содержащие хотя бы одно из следующих слов.** Исключаемые слова заносятся в соответствующее окно ввода.
 - 5) **Язык документа:**
 - любой;
 - русский;
 - английский;
 - украинский.
 - 6) **Формат документа:**
 - любой;
 - HTML;
 - Microsoft Word (.doc);
 - Adobe Acrobat PDF (.pdf).
 - 7) **Дата документа** (в формате «дд/мм/гггг»). Указывается диапазон дат.
 - 8) **Искать документы только на следующих сайтах.** Далее вводится список сайтов через запятую. Примеры: **science.rambler.ru**, **www.lenta.ru**, **www.hosting.ua/~name**. После «/» будут искать только каталоги, начинающиеся с «~»!
- Вывод результатов включает в себя:
- 1) **Сортировку:**
 - Сайты по релевантности;
 - Страницы по релевантности;
 - Страницы по дате (сначала новые);
 - Страницы по дате (сначала старые).

2) **Выдавать:**

- по 15;
- по 30;
- по 50.

3) **Форма вывода.** Можно установить режим «вебмастера», в котором поисковая машина после каждого найденного результата предлагает ссылку на просмотр всех проиндексированных документов с найденного сайта, а также сразу выводит их количество.

<u>ТЕМА 1. Основные понятия сетей.....</u>	<u>3</u>
<u>1. Введение в сети.....</u>	<u>3</u>
<u>2. Принципы построения Internet.....</u>	<u>4</u>
<u>3. Основные сервисы Internet.....</u>	<u>7</u>
<u>1. Основная проблема поиска информации.....</u>	<u>11</u>
<u>2. Методы поиска.....</u>	<u>11</u>
<u>3. Этапы поиска.....</u>	<u>12</u>
<u>ТЕМА 3. Поисковая система Яндекс.....</u>	<u>13</u>
<u>1. Поиск по запросам.....</u>	<u>13</u>
<u>2. Использование формы расширенного поиска.....</u>	<u>19</u>
<u>3. Поиск по каталогу ресурсов.....</u>	<u>20</u>
<u>1. Поиск по запросам.....</u>	<u>21</u>
<u>2. Поиск по каталогу.....</u>	<u>24</u>
<u>3. Использование сервисов АПОРТ.....</u>	<u>27</u>
<u>4. Специальные виды поиска.....</u>	<u>28</u>
<u>ТЕМА 5. Поисковая система GOOGLE.....</u>	<u>29</u>
<u>1. Поиск по запросам.....</u>	<u>29</u>
<u>2. Дополнительные сервисы GOOGLE.....</u>	<u>32</u>
<u>3. Расширенный поиск.....</u>	<u>32</u>
<u>1. Поиск по запросам.....</u>	<u>36</u>
<u>2. Специальные виды поиска.....</u>	<u>39</u>
<u>3. Использование формы расширенного поиска.....</u>	<u>39</u>

Учебное издание

Ибушева Олеся Владимировна

кандидат физико-математических наук

Руженский Олег Владимирович

Саримов Назыф Нуриманович

кандидат физико-математических наук

ИНТЕРНЕТ-ТЕХНОЛОГИИ ПОИСКА И ПОЛУЧЕНИЯ ДЕЛОВОЙ ИНФОРМАЦИИ

ТЕКСТЫ ЛЕКЦИЙ

Корректор Габдурахимова Т.М.

Худ. редактор Федорова Л.Г.

Тех. редактор Горшенин Д.Г.

Сдано в набор 16.03.2010.

Подписано в печать 22.03.2010.

Бумага писчая. Гарнитура Таймс.

Усл. печ. л. 2,6. Тираж 100.

Заказ №8.

НХТИ (филиал) ГОУ ВПО «КГТУ», г. Нижнекамск, 423570,
ул. 30 лет Победы, д. 5а.